

Tilburg University

Exploring the effect of a middle response category on response style in attitude measurement

Moors, G.B.D.

Published in:
Quality & Quantity

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Moors, G. B. D. (2007). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, 42(6), 779-794.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Exploring the effect of a middle response category on response style in attitude measurement

Guy Moors

© Springer Science+Business Media B.V. 2007

Abstract In this research we focus on the link between response style behaviour in answering rating data such as Likert scales and the number of response categories that is offered. In a split-ballot experiment two versions of a questionnaire were randomly administered. The questionnaires only differed in the number of response categories, i.e. 5 vs. 6 categories. In both samples a latent-class confirmatory factor analysis revealed an extreme response style factor. The 6-response categories version, however, revealed the more consistent set of effects. As far as the content latent-class factors, i.e. familistic values and ethnocentrism, are concerned, results were fairly similar. However, a somewhat deviant pattern regarding the familistic values items in the 6-response categories version suggested that this set of items is less homogeneous than the set of ethnocentric items. The effect of gender, age and education was also tested and revealed similarities as well as differences between the two samples.

Keywords Response style · Attitudes · Number of response categories · Survey research · Latent class factor analysis

1 Introduction

In survey research attitudes are often measured by sets of items with similar response scales indicating the level of agreement with these items, e.g. Likert scales. In general, this type of questions is referred to as rating (or agreement) questions as opposed to ranking (or preference) questions in which respondents compare items and choose among them. Undoubtedly, the popularity of agreement scales has much to do with the fact that they are fairly easy to administer and that a multitude of methods can be used to model this type of data. However, there is a growing awareness among

G. Moors (✉)

Department of Methodology and Statistics, University of Tilburg, FSW-MTO Room S110,
Warandelaan 2, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: guy.moors@uvt.nl

survey researchers that rating questions may be vulnerable to response style behaviour causing non-random response error. In an ideal world respondents answer to a set of items simply and solely on a content basis. In real life, however, other characteristics of respondents may affect the way they answer to questions. Two such examples for which we observe a kind of revival in interest among social researchers (Cheung and Rensvold 2000; Billiet and McClendon 2000; Moors 2003) are acquiescence and extreme responses. Acquiescence refers to the tendency to agree with issues irrespective of the content of these issues, whereas an extreme response style is adapted when respondents tend to pick the extremes of a response scale. A question that is rarely raised in the literature on response styles is to what extent question format has an effect on the likelihood of response bias. One of few exceptions is the suggestion that ranking or forced choice questions prevent respondents of applying an acquiescence or extreme response style (Berkowitz and Wolkon 1964; Shuman and Presser 1981; Toner 1987). That question format has an impact for the survey response process, however, has been repeatedly argued in the literature (van der Veld and Saris 2005). In this paper, we focus on one particular issue, i.e. the relationship between (extreme) response style and offering a middle 'neutral' position in attitude questions. We explain the conceptual rational of the model that is used to identify response style, i.e. a latent class (LC) confirmatory factor model of two sets of balanced questions in which one LC-factor identifies an extreme response style and two LC-factors measure the two content factors. In a split-ballot design we compare response scales with 5- and 6-response categories. Before discussing the findings from this research we present a short overview of perspectives on response styles, and relate it to the issue of response categories.

2 Perspectives on response styles

Perspectives on the concept of response style differ and range from minimizing—or even neglecting—the problem to highlighting it. Among the latter are those who argue that response styles are meaningful personality constructs, and as such are potentially measurable (e.g. Couch and Keniston 1960; Jackson and Messick 1965). From this perspective it is not justified to refer to the phenomenon as a response 'style' but as a response 'trait' that is not intended to be measured by items of a particular measurement instrument. The second perspective that recognizes the concept of response style only thinks of it as a statistical nuisance that should be controlled for in empirical models. As Mellenbergh (2001) has pointed out response bias arises in a situation in which a person responds to item properties other than the location of that item in the construct to be measured. In this second perspective response style is an artefact in measurement that depends on characteristics of the instrument that is used. Finally, some researchers question the meaningfulness of the concept of response bias altogether. Nunnally (1978) for instance, argues that there is little evidence that agreement tendency is an important measure of personality. He also claims that agreement bias is not an important source of systematic invalidity in measuring constructs. Rorer also questions the relevance of the concept by referring to 'the great response-style myth' (1965). Without going into detail, Rorer's arguments also relate to the lack of consistent empirical evidence, a finding that is recently confirmed by Ferrando et al. (2004) when they researched convergent validity of acquiescence in balanced and separate acquiescence scales. Nevertheless,

in our opinion one should be prudent to generalize the findings from these studies. These findings do not suggest that one may feel confident in ignoring possible bias due to response style in survey research. After all, regardless whether one ‘believes’ that response style is a personality construct, a statistical nuisance, or even a non-significant factor, it is important to rule out that response style biases the measurement of constructs or implies misspecification of the relationship of content traits with covariates. Already in 1981, in their standard work on question form, wording and context in attitude surveys, Schuman and Presser warned that even the ‘dismissal of the importance of acquiescence in psychological investigations is not incompatible with the assumption of survey researchers that acquiescence is quite important in survey data’ (pp. 205–206). They refer to survey-based interpretations of acquiescence that hypothesize an inverse association with education. Furthermore, response styles may be greatest when vague, ambiguous or difficult to answer items are involved. From political science research, for instance, we have learned that the political value structure of respondents involved in politics can differ from general public (e.g. Inglehart 1990, chapter 9). Hence, it remains important to account for possible response bias in measuring constructs and estimating the effect of covariates.

3 Modelling response styles

It does not come as a surprise, given the aforementioned controversies regarding response style, that there is no single accepted statistical procedure to take response bias into account in attitude survey (Cheung and Rensvold 2000). However, within the context of structural equation modelling an approach has been suggested that is promising (Billiet and McClendon 2000; Moors 2003). Promising in the sense that the approach does not explicitly ally with a particular perspective on response styles as discussed in the previous section. Billiet and McClendon (2000) define the basic premises of this approach in the context of acquiescence response style. Their key argument is that it is possible to model a response style factor by means of a structural equation modelling of two balanced sets of items measuring two different constructs. As such, they combine the two major procedures for assessing acquiescence (Ferrando et al. 2004), i.e. (a) the balanced scale procedure and (b) the separate acquiescence scale approach. A balanced scale includes items that are negatively as well as positively worded and that are supposed to measure the same latent construct. It is assumed that acquiescence to the items in the positive direction is cancelled out by acquiescence to the items in the negative direction. One major critique, however, is that it is fairly difficult to operationalize complete reversals of items and hence that positively and negatively worded items may measure different or only partly related constructs. In Billiet and McClendon’s procedure this critique is overcome by combining balanced items with the idea that items should be pooled from heterogeneous sets. Pooling items from heterogeneous sets allows operationalizing a separate acquiescence scale. Hence, Billiet and McClendon estimate a confirmatory three-factor model including two so-called ‘content’ factors and one so-called ‘style’ factor. To identify the latter style factor they impose equality restrictions on the factor weights of all items of the two sets of questions. Hence, regardless of the content of the items the same factor loading is imposed on all positive and negative issues of two sets of items. Obviously, this style factor highly correlates with an index that counts the number of ‘completely agree’ answers on all items of the two sets of questions. Moors (2003, 2004) adapted

Table 1 Crosstabulation of two ‘ethnocentric’ issues, adjusted residuals

		Cultural threat				
		1 c.d.	2 d.	3 n.	4 a.	5 c.a.
Job threat	1 c.d.	15.2	2.8	−4.6	−5.6	−3.7
	2 d.	−3.2	8.2	1.4	−3.2	−6.0
	3 n.	−4.4	−4.8	5.5	2.5	−0.4
	4 a.	−4.1	−5.9	−2.4	6.4	6.4
	5 c.a.	0.0	−3.1	−3.5	−1.1	10.4

c.d. completely disagree, *d.* disagree, *n.* neutral, *a.* agree, *c.a.* completely agree

the same rational within the context of a LC confirmatory factor approach to model extreme response bias. This approach is also adopted in this research. Conceptually a latent-class factor approach for analysing the latent structure of categorical variables is similar to the confirmatory factor (Lisrel) approach for the analysis of continuous variables. Explaining the technical details of the method is beyond the purpose of this paper. For the reader who is interested we refer to the following references that provide a more in-depth reading on the subject: Heinen (1996), Vermunt (1997) and Magidson and Vermunt (2001). However, we need to point out the advantage of the LC-factor approach adopted in this paper compared to the more commonly known confirmatory factor approach with Lisrel. The major difference is that a LC-factor approach does not use a correlation or variance/covariance matrix as an input. Instead it analyses the underlying pattern in the cross-classification of the responses pertaining to the manifest variables of interest. For example, a LC-factor model including four items, each of which has five response categories, involves analysing a ‘ $5 \times 5 \times 5 \times 5$ ’ table. Analysing such a cross-classification is perhaps more ‘complex’ to understand, but at the same time more informative as is demonstrated in Table 1.

Table 1 represents adjusted residuals of the crosstabulation of two ethnocentric-worded questions that are used in this research (cfr. *infra*). As is clear from the table, a positive association between the two items is observed. Such a relationship can be represented by one ordinal (gamma) or one linear (Pearson) correlation. The information that is ignored with these two summary measures, however, is that the residuals do not continuously decrease the further one moves from the main diagonal. For instance, respondents who fully agree with the ‘cultural threat’ issue are less likely to ‘disagree’ than to ‘completely disagree’ with the issue of ‘job threat’. This pattern suggests that an extreme response style might have influenced the way respondents answered to the questions (Moors 2003). A LC-factor approach with nominal indicators uses the full information from multiple frequency tables. By consequence, an effect of a LC-factor on a response variable is represented by several coefficients, i.e. equal to the number of response categories. Hence, if 5-response categories are offered, five coefficients are necessary to estimate the effect of one LC-factor on one item. Such a model is less parsimonious, but—as will be demonstrated in this paper—is flexible in diagnosing response styles.

4 Number of response categories, the ‘middle’ option and response style

Research on number of response categories has primarily focused on the reliability of attitude scales. Most research on this topic can be found within the psychological

and psychometric literature; in the context of survey research this topic is less covered (Alwin 1992). Information theorists would probably argue that this question is trivial since the more response categories that are presented the more bits of information are conveyed. In survey research, however, the key question is whether there are an optimal number of response categories. This question is raised, not only from a statistical point of view, but also from a cognitive point of view (Alwin 1992). After all, too many categories may go beyond a respondent's ability to distinguish among categories. Hence, recent research has focused attention to detecting a sufficient number of response categories that optimizes reliability and at the same time does not cause unnecessary burden upon a respondent (Viswanathan et al. 2004). The 'ideal' number of response categories, however, has not (yet) been established. Presumably, this 'ideal' number may also depend on the nature of attitude questions that is asked. Nevertheless, Alwin (1992) has suggested that this number might range between 4 and 7 categories with little left to gain in increasing the number higher than 7.

The question of offering a middle 'neutral' response alternative—which is the central focus of our research—is related to the aforementioned issue regarding the number of response categories in attitude research. Previous research has demonstrated that when this middle option is offered, it is far more likely to be chosen. Furthermore, it is argued that people who select a middle response alternative do not necessarily answer the question in the same way as other respondents if forced to choose sides on the issue (Bishop 1987; Kalton et al. 1980). The discussion about offering a middle response category may also extend to the validity of a measurement model. Hurley (1998) has argued that a mild response style can be regarded as the counterpart of the extreme response style. A mild response style implies a tendency to use the middle categories, while avoiding the extremes of a scale. As Presser and Schuman (1980) argued, less intense respondents are more affected by the presence or absence of a middle response category than respondents that feel strongly about the attitude. To some extent the mild response style is the conceptual counterpart of extreme response behaviour. In this research, we experimented with 5- vs. 6-response categories. Given the previous arguments we expect that offering 6-response categories will more easily allow to observe a mild response style as the counterpart of an extreme response style.

5 Data and methodology

A short questionnaire was designed and administered in the context of the CentER-panel web-survey that was established in 1991. The random sample consists of over 2,000 households in The Netherlands representative of the Dutch population. On a weekly basis panel members complete a questionnaire on the Internet from their home. Households without Internet access are supplied with a set-top box with which questionnaires can be completed using a television screen as a monitor. A benefit of the web-based design is that it is easy to implement a split-ballot design in which respondents are randomly assigned to answer one of two versions of the questionnaire. The same questionnaire was administered only changing the number of response categories, i.e. 5 vs. 6. Including our split-ballot experiment in a representative sample survey also increases the validity of our findings compared to other research that often uses homogeneous student populations to conduct this type of experimenting.

Table 2 Gender, age and education by questionnaire format

		Version		Total		
		1 = 5 cat.	2 = 6 cat.			
Gender	Man	51.6%	53.8%	1,083	Pearson chi-Square <i>p</i> -value	1.001
	Woman	48.4%	46.2%	971		0.317
Ages	15–24 years	5.8%	7.0%	132	Pearson chi-Square <i>p</i> -value	18.455
	25–34 years	17.8%	16.4%	351		
	35–44 years	15.6%	20.2%	368		
	45–54 years	20.6%	22.9%	447		
	55–64 years	18.3%	17.2%	364		
	65 years and older	22.1%	16.3%	392		
Education	Primary education	6.1%	5.8%	122	Pearson chi-Square <i>p</i> -value	3.909
	Pre-vocational education	27.3%	28.3%	570		
	Pre-university education	13.3%	13.9%	279		
	Senior vocational colleges	18.0%	19.3%	383		
	Vocational colleges	24.0%	23.8%	490		
Total	University education	11.4%	8.9%	208	<i>p</i> -value	0.563
		1,002	1,052	2,052		

In Table 2, we compare the two samples with regard to the main covariates that are used in this research.

From Table 2 we read that the two samples are fairly comparable. The age distribution, however, slightly differs especially at the oldest age category. Differences between the other age categories proved not to be significant.

The questionnaire included two sets of four questions that were intended to measure ‘familistic attitude’, i.e. attitudes towards family and children, and ‘ethnocentrism’, i.e. attitudes towards immigrants. Each set included two positively (+) and two negatively (–) worded items. An overview of these items is presented in Table 3.

The response scales that were presented to the respondents were fully labelled, distinguishing between ‘completely disagree’, ‘disagree’, ‘agree’ and ‘completely agree’ in both versions. A ‘neutral’ category was included in the 5-categories version; this was substituted by ‘rather disagree’ and ‘rather agree’ in the 6-categories version.

Latent Gold 4.0 was used to estimate a confirmatory LC-factor model with one ‘style’ factor influencing all eight items, one ‘content’ factor influencing the responses on the four ‘familistic’ attitudes, and a second ‘content’ factor that influences the responses to the four ‘ethnocentric’ attitudes. The two content factors are allowed to correlate. This type of analysis in which the manifest items are treated as nominal response variables, and the LC-factor as a discrete interval variable is referred to as a latent trait approach. For ease of expose, assume a model with two sets of items (A and B) and two LC-factors (X). Then the conditional response probabilities of this latent-class factor model can be written as:

$$\prod_{k=1}^4 \pi(A_1 A_2 B_1 B_2 | X_1 X_2). \quad (1)$$

Table 3 Overview of items measuring familistic and ethnocentric attitudes

(a) Familistic attitudes (adapted from the European Values Surveys)		
a1. A working mother can establish just as warm and secure a relationship with her children as a mother who does not work. [WORKING MOTHER]	(-)	
a2. A pre-school child is likely to suffer if his or her mother works. [PRE-SCHOOL CHILD]	(+)	
a3. A job is alright but what most women really want is a home and children. [FAMILY ORIENTATION]	(+)	
a4. There is more in life than family and children, what a woman also needs is a job that gives her satisfaction. [JOB ORIENTATION]	(-)	
Note: (+) refers to familistic attitudes; (-) refers to emancipated attitudes		
(b) Ethnocentric attitudes (adapted from the Belgian 1995 General Elections Survey)		
b1. In general, immigrants can be trusted. [TRUST]	(-)	
b2. Guest workers endanger the employment of persons who are born in the Netherlands. [JOB THREAT]	(+)	
b3. The presence of different cultures enriches Dutch society. [CULTURAL ENRICHMENT]	(-)	
b4. Muslims are a threat for Dutch culture and our values. [CULTURAL THREAT]	(+)	
Note: (+) refers to ethnocentrism; (-) refers to tolerant attitudes towards immigrants		

The response probabilities of this model are restricted by means of logit models with linear terms:

$$\eta_{A_1 A_2 B_1 B_2 | X_1 X_2} = \beta_{A_1}^0 + \beta_{A_2}^0 + \beta_{B_1}^0 + \beta_{B_2}^0 + \beta_{A_1}^1 \cdot v_{X_1} + \beta_{A_2}^1 \cdot v_{X_1} + \beta_{B_1}^1 \cdot v_{X_2} + \beta_{B_2}^1 \cdot v_{X_2}. \quad (2)$$

Since a LC-factor approach assumes that the factors are discrete interval (or ordinal) variables, the two-variable terms (e.g. $\beta_{A_1}^1 \cdot v_{X_1}$) are restricted by using fixed category scores for the different categories of the LC-factor. Equidistant scores v_X range from 0 to 1, with the first category of a factor getting the score 0 and the last category the score 1. Hence, a LC-factor with, for instance, four categories gets the scores 0, 1/3, 2/3 and 1. As such the categories of the factors are ordered by the use of fixed equal-interval category scores. The β 's indicate the strength of the relationship between factors and response variables. Equation (2) identifies a confirmatory LC-factor model with factor X_1 influencing the response probabilities of items A_1 and A_2 , and factor X_2 influencing items B_1 and B_2 .

In preparation of this paper, we estimated several models that differed in the number of categories of the LC-factors. We tested models with two, three and four categories. Results of the two-category model differed slightly from the models in which LC-factors were identified with three and four categories. Results of the three and four categories models were highly similar. The analyses reported in this paper identified LC-factors with four ordered categories. The covariates that are listed in Table 2 are also included in the analysis. After all, we have argued that response styles may bias the measurement of constructs as well as the effect of covariates. In the next section, we first discuss the comparison of the measurement models of the two samples. Then we focus on covariate effects.

6 Exploring the effect of a ‘neutral’ response category on the measurement of ‘style’ and ‘content’ factors

There are different possibilities to present the LC-factor results. One obvious choice would be to present the β 's as defined in Eq. 2. Since our analysis includes 5 or 6 coefficients for each item per factor, this would result in a huge table, which would be difficult to interpret unless one is familiar with the method. Furthermore, the β 's do not have an upper limit which makes them more difficult to interpret and compare. For these reasons we have opted for a graphical presentation of the probability means (Figs. 1, 2). A probability mean is the mean latent-class factor score for each response category and ranges from 0 to 1. The order of the response categories of the negatively worded item has been reversed in Figs. 1 and 2 to be consistent in content to the positively worded items. Hence, the order for negatively worded items ranges from ‘completely agree’ (c.a.) to ‘completely disagree’ (c.d.), whereas the positively worded items range from ‘completely disagree’ (c.d.) to ‘completely agree’ (c.a.). Figure 1 compares the results of the effect of the extreme ‘style’ LC-factor on the eight items. Figure 2 includes the comparison of the two ‘content’ LC-factors, i.e. ‘familistic values’ and ‘ethnocentrism’.

The main finding from Fig. 1 is that regardless of the number of response categories that is used, an extreme response style is clearly observed. The U-shape curve indicates that the mean probability is highest among the two extreme response categories ‘completely agree’ and ‘completely disagree’. The principal difference between the two question formats is how the intermediate response categories position. When 5-response categories are offered, there is little difference between the ‘neutral’ category on the one hand, and the ‘agree’ or ‘disagree’ categories on the other. There are three exceptions in which the mean probability for ‘disagree’ is (somewhat) higher, i.e. ‘working mother’, ‘job orientation’ and ‘trust’. It is difficult to pinpoint the rational of these exceptions. Nevertheless, what can be concluded from comparing the intermediate response categories in Fig. 1(a) is that a ‘modest’ response style is not unequivocally observed—assuming of course, that such a modest response style leads to a greater prevalence of a neutral response. In Fig. 1(b) in which the results of the 6-response categories sample are summarized, a more consistent picture emerges. First of all, it confirms the findings from Fig. 1(a) that an extreme response style is clearly observed. Second, one particular response category clearly positions between the two extreme response on the one hand, and the other response categories on the other hand. This intermediate response category is not related to a particular response number and label. Rather it corresponds with ‘disagree’ for the negatively worded questions and ‘agree’ for the positively worded questions. Hence, it is in consistency with the content of the question, i.e. reflecting higher agreement with ‘familistic values’ and ‘ethnocentrism’. Only as far as the ‘cultural threat’ issue is concerned, this is less strongly observed. There is also a slight tendency for the two middle response categories to have the lowest mean probability values. However, to conclude from this that a ‘modest’ response style is emerging is difficult. For this, we need a more clear-cut difference of the two middle response scales with both adjacent ‘agree’ and ‘disagree’ categories. In this study, this is only the case with one of these adjacent categories. However, this is not to say that a ‘modest’ response style is non-existing either. After all, it is possible that lower agreement with ‘familistic’ and ‘ethnocentric’ issues may reflect a ‘modest’ response style. Future exploration of the issue by increasing the number of response categories may provide some answer.

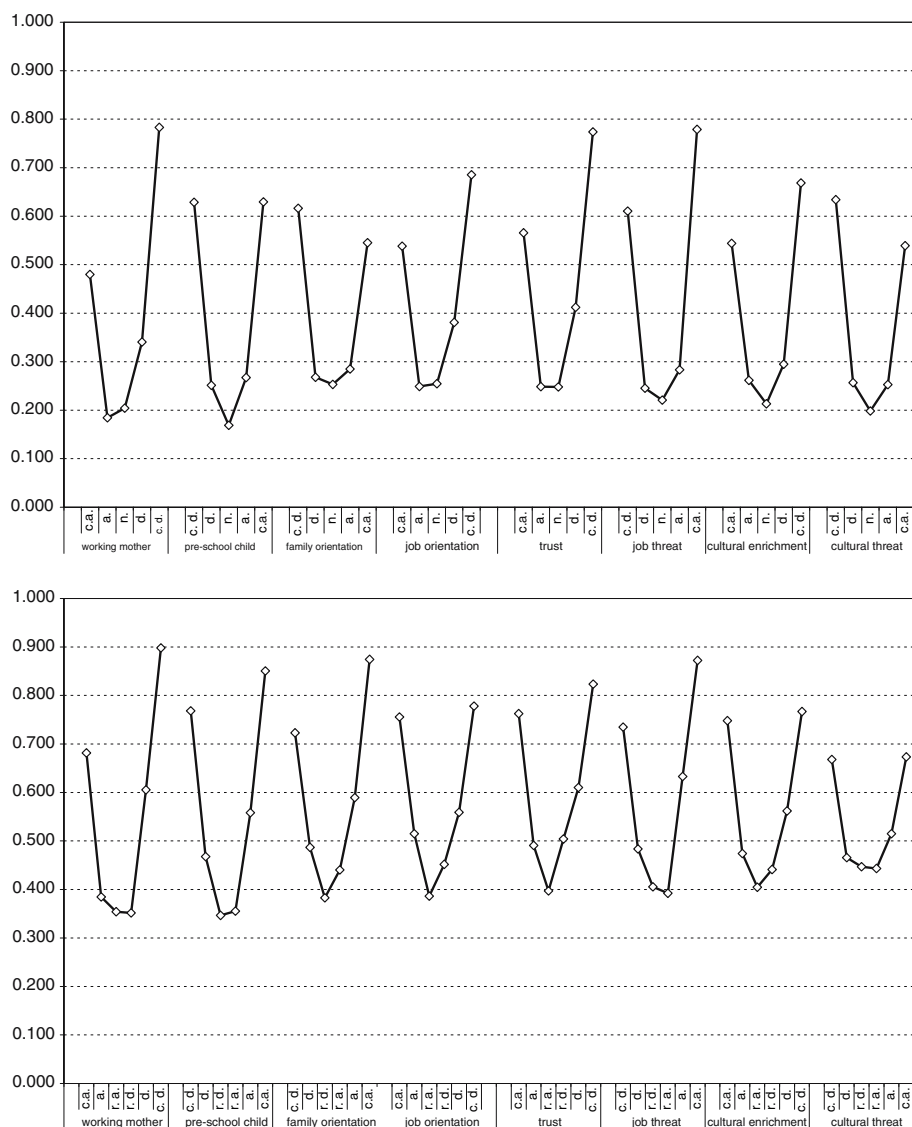


Fig. 1 Mean probability scores on LC Factor 1 'extreme response style'. (a) Sample 1: Five response categories. (b) Sample 2: Six response categories

The two 'content' factors are reproduced by the second and third LC-factor of the model. The general finding is that moving from left to right in response categories presented in Fig. 2, the mean probability scores of the LC-factors also increase, which is in consistency with the content of each item. There are, however, some differences that need a closer look. As far as the 'familistic values' LC-factor is concerned, there is a clear difference between the 5- and 6-response categories models. The difference is that in the 6-response categories model the mean probability score of the 'job satisfaction' issue first increases as the level of disagreement increases, but this is reversed at

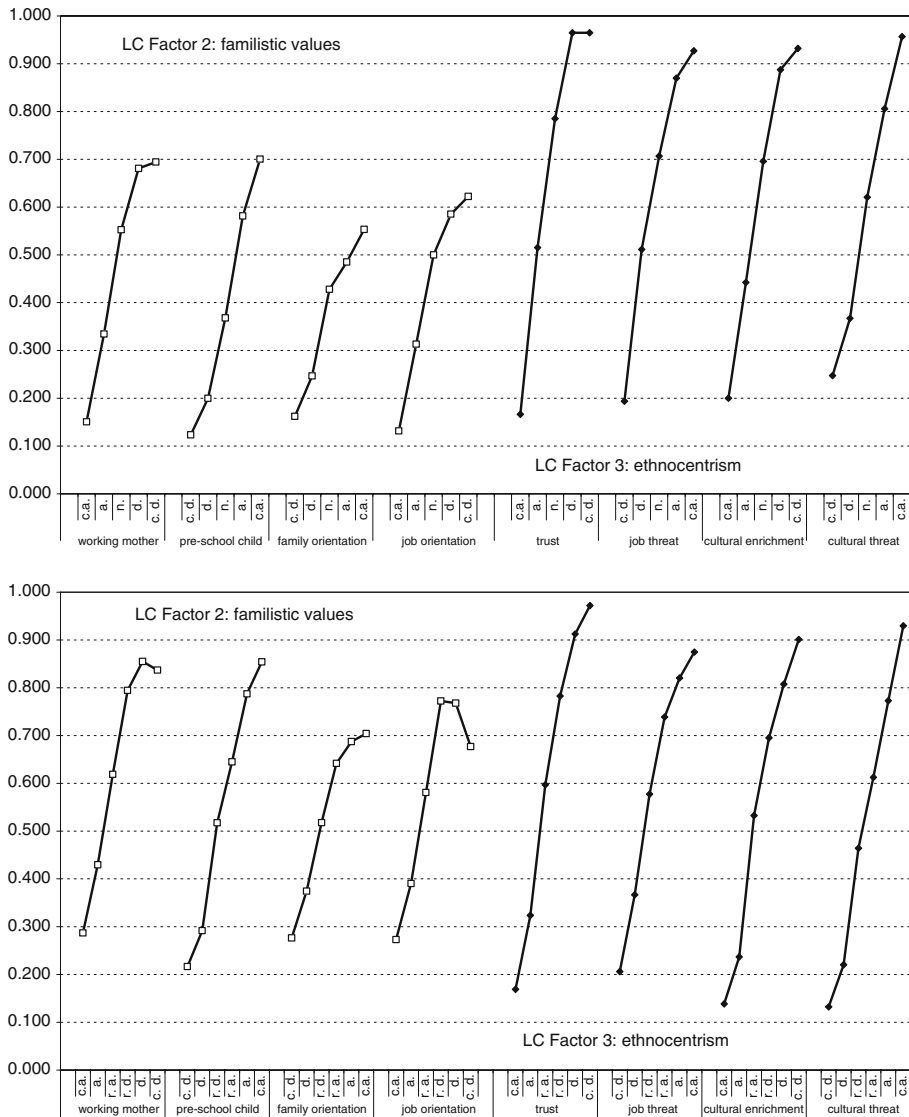


Fig. 2 Mean probability scores on LC Factor 2 ‘familistic values’ and LC Factor 3 ‘ethnocentrism’. (a) Sample 1: Five response categories. (b) Sample 2: Six response categories

the end of the scale. The mean probability score of the ‘completely disagree’ category is lower than the score of the ‘disagree’ category. The same pattern is observed for the first issue ‘working mother’, but in this case the differences are very small. A likely explanation of this finding could be that the set of ‘familistic issues’ is less homogeneous than the set ‘ethnocentric’ issues. Taken at face value the ‘working mother’ and ‘pre-school child’ issues refer to the impact of a working mother on children; the other two issues refer to the conflicting roles of family versus job for women. Furthermore, inspecting the pseudo- R^2 values (Table 4) for each adds support to this

Table 4 Pseudo R^2 (explained variance by the LC-factors)

	Sample 1 5 cat.	Sample 2 6 cat.
Working mother	0.322	0.283
Pre-school child	0.297	0.277
Family orientation	0.089	0.090
Job orientation	0.145	0.160
Trust	0.273	0.255
Job threat	0.238	0.178
Cultural enrichment	0.236	0.234
Cultural threat	0.262	0.216

interpretation. These pseudo- R^2 values indicate the ‘explained variance’ of the items by the LC-Factors. The explained variance of the two items of the familistic values dimension, i.e. ‘family orientation’ and ‘job orientation’, is lower than the explained variance of the other items in the analyses. This was true for both versions of the questionnaire. The fact that the more detailed six categories scale revealed a non-ordinal relationship of ‘job orientation’ with the latent content factor of familistic values might merely point to this issue.

There is little difference in how the two versions of the ‘ethnocentric’ items relate to the corresponding LC-factor, i.e. probability means increase in consistency with the ordered response categories. An interesting observation is that in the model with 5 response categories there is a large gap between the first and second response category, after which differences between adjacent categories decrease. In the model with 6 response categories, the adjacent categories are more evenly spread.

7 Comparing the effect of covariates

From the previous section we have learned that the measurement models of the 5- and 6-response scales were similar and at the same time revealed particularities. The next step is to compare the two models as far as the effect of selected covariates are concerned. In Table 5, we present the effect of gender, age and education on the extreme response style factor (LC-factor 1), the familistic orientation (LC-factor 2) and ethnocentrism (LC-factor 3). Recall that the figures presented in Table 5 are the structural part of the LC structural equation model that was also used to identify the LC-factors. Drawing an analogy with Lisrel modelling, Hagenaars (1990, see also: Goodman 1974) has referred to this model as a ‘modified Lisrel approach’. The three covariates are treated as categorical. Deviation coding was used which means that the overall effect (beta) of each covariate is fixed to zero, which is the reference value. Associated standard errors are presented, as well as the p -value of the Wald-statistic which indicates an overall significance of a covariate.

A first general finding is that the extreme response LC-factor does not significantly relate to any of the three covariates, suggesting that this response style may reflect a personality characteristic which has little to do with social background. In both models, gender is related to the second LC-factor ‘familistic values’ with men being more ‘conservative’ than women. The education variable was not significant in the 5-response categories analysis although the beta’s suggest that ‘familistic values’ decrease with

Table 5 Effect of covariates on the LC-factors

Covariates	LC-factor1 extreme response			LC-factor2 familistic values			LC-factor3 ethnocentrism		
	Beta	s.e.	Wald <i>p</i> -value	Beta	s.e.	Wald <i>p</i> -value	Beta	s.e.	Wald <i>p</i> -value
(a) Sample 2: 5 response categories									
Gender									
Man	0.112	0.175	0.52	0.702	0.247	0.00	-0.038	0.162	0.82
Woman	-0.112	0.175		-0.702	0.247		-0.038	0.162	
Age groups									
15–24 years	0.278	0.503	0.37	-0.618	0.678	0.18	0.843	0.613	0.00
25–34 years	-0.047	0.311		-1.128	0.562		0.951	0.376	
35–44 years	0.046	0.314		0.321	0.438		-0.395	0.360	
45–54 years	0.415	0.276		0.248	0.389		-0.525	0.328	
55–64 years	-0.154	0.300		0.840	0.436		-1.313	0.352	
65 years and older	-0.537	0.303		0.337	0.385		0.439	0.337	
Educational level									
Primary education	-0.253	0.510	0.26	0.959	0.676	0.27	1.654	0.643	0.00
Pre-vocational education	0.201	0.334		0.575	0.384		0.989	0.327	
Pre-university education	-0.693	0.371		0.366	0.473		-0.743	0.384	
Senior vocational colleges	-0.115	0.308		-0.337	0.411		0.824	0.360	
Vocational colleges	0.373	0.276		-0.483	0.414		-1.251	0.317	
University education	0.487	0.380		-1.081	0.597		-1.472	0.420	

Table 5 continued

Covariates	LC-factor1 extreme response			LC-factor2 familistic values			LC-factor3 ethnocentrism		
	Beta	s.e.	Wald <i>p</i> -value	Beta	s.e.	Wald <i>p</i> -value	Beta	s.e.	Wald <i>p</i> -value
(a) Sample 2: 6 response categories									
Gender									
Man	0.163	0.139	0.24	0.691	0.156	0.00	0.055	0.137	0.69
Woman	-0.163	0.139		-0.691	0.156		-0.055	0.137	
Age groups									
15–24 years	-0.422	0.421	0.58	0.332	0.479	0.31	0.987	0.441	0.11
25–34 years	-0.302	0.306		-0.231	0.348		0.242	0.303	
35–44 years	-0.107	0.269		-0.165	0.308		-0.540	0.271	
45–54 years	0.188	0.263		-0.348	0.294		-0.306	0.259	
55–64 years	0.424	0.307		0.295	0.332		-0.072	0.294	
65 years and older	0.220	0.318		0.706	0.353		-0.311	0.301	
Educational level									
Primary education	1.007	0.585	0.30	0.040	0.527	0.00	0.582	0.468	0.00
Pre-vocational education	-0.598	0.286		1.154	0.306		0.917	0.270	
Pre-university education	-0.301	0.321		-0.365	0.358		0.106	0.313	
Senior vocational colleges	-0.115	0.298		0.298	0.325		0.470	0.289	
Vocational colleges	0.020	0.278		-0.093	0.299		-1.127	0.266	
University education	0.027	0.387		-1.112	0.465		-0.948	0.386	

educational level. In the model with 6-response categories, education was significant. The two categories of education that stand out are (a) the highest level of university education ($\beta = -1.112$) that is the least 'conservative' in 'familistic values' and (b) the lower category of pre-vocational education ($\beta = +1.154$) that is the most conservative. Age differences in 'familistic values' were not significant.

As far as the third 'ethnocentrism' LC-factor is concerned education was the single most important covariate in both samples. Lower levels of education showed the highest level of ethnocentrism, whereas the higher educated were the least ethnocentric. The 5-response categories model indicates that ethnocentrism decreases with age, except for the oldest category. Recall that this effect is controlled for educational differences. In the 6-response categories model the overall effect of age is not significant, although the youngest category has also a higher level of ethnocentrism similar to the 5-categories sample.

8 Conclusion and discussion

The purpose of this research was to explore the relationship between an extreme response style and the number of response categories that is offered with Likert-type of questions. We compared two versions of the same questionnaire offering, respectively, 5- and 6-response categories. Results indicated similarities of the two models as well as some striking differences.

The two versions of the questionnaire clearly revealed an extreme response style as indicated by the highest probability among the two extremes of each item included in the analyses. An advantage of the model with 6 response categories was that agreement with positively worded items and disagreement with negatively worded items consistently fell in between the high probability of the extreme response categories and the low probability of the remaining categories. To some extent this indicates that the middle categories may function as the mirror image of the extreme response, i.e. a mild response style. However, a more decisive conclusion needs additional research in which more response categories are offered than the two versions that were administered in this research.

As far as the two content factors is concerned, the general tendency in each model is that the mean probabilities LC-factor scores increases consistently with the ordered categories of the response items. However, this was less clearly the case with two items of the set familistic attitudes when 6 response categories were offered. A hypothetical interpretation is that the set of four familistic attitudes is less homogeneous in content than the set ethnocentric attitudes. The more response categories that is offered, the more likely the heterogeneity of the items may be revealed. Hence, the difference between the 5- and 6-response categories model. Supportive evidence for this argument about a lesser degree of homogeneity among the familistic items was found in the lower explained variance of two items from this set. However, what cannot be excluded is that perhaps respondents need less categories to express their opinion about familistic attitudes. This reflection is guided by the literature on the 'optimal' number of categories that is necessary to express one's opinion. Future research should clarify this issue. Again, our research needs to be extended by increasing the number of response categories. Finally, there was less difference between the 5- and 6-response categories models as far as the ethnocentric items is concerned, although

the 6-responses model revealed more equidistant differences in LF-factor values than the 5-responses model.

The effect of age, gender and education in the two models is also compared. Covariates were not significantly related to the response style LC-factor, which was partly surprising given the suggestion in the literature that response biases may be highest among the less educated. On the other hand, to the extent that an extreme response style may be regarded as a personality trait, this finding is perhaps less surprising. Gender proved to be the single most important covariate in explaining familistic values, with men being the more conservative category. Education was significant in the 6-response model, but not in the 5-response model. Of course, this difference could be attributed to the fact that the measurement model of the second 'familistic values' LC-factor is different in these two models. On the other hand, in both models respondents with a university degree were the least likely to hold traditional 'familistic values', and this contrasts with lower education. But there are differences that remain unexplained until we resolve the issue of the measurement of 'familistic values'. Increasing the number of items to measure this dimension and to research the level of homogeneity in content of these items is necessary.

Like any exploratory research, this research reveals a number of findings and at the same time raises some question. In this discussion, we already referred to some issues that needs further attention in future research on the effect of the number of response categories in models with response styles, i.e. increasing the number of response categories; questioning the optimal number of response in relationship to the content of the items; and selecting homogeneous sets of items. In this research, we have explored the relationship of response style and number of response categories in a split-ballot design. This is a nice design to explore the issue, but it is not an ideal design in helping to decide on the 'best' possible response format. For this reason, our final suggestion is to explore the aforementioned issues more in-depth by developing MTMM designs (Saris et al. 2004) that are more powerful in making suggestions about an 'appropriate' question format. We do need to keep in mind, however, that there may be different 'optimal' solutions, depending on the content of the items that are researched.

Acknowledgements The author gratefully acknowledges CentREdata and its director, Marcel Das, for including the split-ballot experiment in their web-panel survey.

References

- Alwin, D.F.: Information transmission in the survey interview: number of response categories and the reliability of attitude measurement. *Sociol. Methodol.* **22**, 83–118 (1992)
- Bishop, G.F.: Experiments with the middle response alternative in survey questions. *The Public Opin. Q.* **51**, 220–232 (1987)
- Berkowitz, N.H., Wolkon, G.H.: A forced-choice form of the F-scale free of acquiescent response set. *Sociometry* **24**, 54–56 (1964)
- Billiet, J.B., McClendon, M.J.: Modeling acquiescence in measurement models for two balanced sets of items. *Struct. Equation Model.* **7**, 608–628 (2000)
- Cheung, G.W., Rensvold, R.B.: Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *J. Cross-Cult. Psychol.* **31**, 187–212 (2000)
- Couch, A., Keniston, K.: Yeasayers and Naysayers: agreeing response set as a personality variable. *J. Abnorm. Soc. Psychol.* **60**, 151–74 (1960)
- Ferrando, P.E., Condon, L., Chico, E.: The convergent validity of acquiescence: an empirical study relating balanced scales and separate acquiescence scales. *Pers. Individ. Dif.* **37**, 1331–1340 (2004)

- Goodman, L.A.: The analysis of systems of qualitative variables when some of the variables are unobservable. Part I—a modified latent structure approach. *Am. J. Sociol.* **79**, 1179–1259 (1974)
- Hagenaars, J.A.: *Categorical Longitudinal Data—Loglinear Analysis of Panel, Trend and Cohort Data*. Newbury Park, Sage (1990)
- Heinen, T.: *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Sage Publications, Thousand Oaks, CA (1996)
- Hurley, J.R.: Timidity as a response style to psychological questionnaires. *J. Psychol.* **132**, 201–210 (1998)
- Inglehart, R.: *Culture Shift in Advanced Industrial Society*. Princeton University Press, Princeton (1990)
- Jackson, D.N., Messick, S.J.: Acquiescence: the nonvanishing variance component. *Am. Psychol.* **20**, 498 (1965)
- Katlon, G., Roberts, J., Holt, D.: The effects of offering a middle response option with opinion questions. *Statistician.* **29**, 65–78 (1980)
- Magidson, J., Vermunt, J.K.: Latent class factor and cluster models, bi-plots, and related graphical displays. *Sociol. Methodol.* **31**, 223–264 (2001)
- Mellenbergh, G.J.: Outline of a faceted theory of item response data. In: Boomsma, A., van Duijn, M.A.J., Snijders, T.A.B. (eds.) *Essays on Item Response Theory*, pp. 415–432. Springer, Berlin Heidelberg, New York (2001)
- Moors, G.: Diagnosing response style behavior by means of a latent-class factor approach. Socio-demographic correlates of gender role attitudes and perceptions of ethnic discrimination reexamined. *Qual. Quant.* **37**, 277–302 (2003)
- Moors, G.: Facts and artifacts in the comparison of attitudes among ethnic minorities. A multigroup latent class structure model with adjustment for response style behavior. *Eur. Sociol. Rev.* **20**, 303–320 (2004)
- Nunnally, J.C.: *Psychometric Theory*. McGraw Hill, New York (1978)
- Presser, S., Schuman, H.: The Measurement of a Middle Position in Attitude Surveys. *Public. Opin. Q.* **44**, 70–85
- Rorer, L.G.: The great response-style myth. *Psychol. Bull.* **63**, 129–156 (1965)
- Saris, W.E., Satorra, A., Coenders, G.: A new approach to evaluating the quality of measurement instruments: the split-ballot MTMM design. *Sociol. Methodol.* **34**, 311–347 (2004)
- Shuman, H., Presser, S.: *Questions and Answers in Attitude Surveys*. Academic, New York (1981)
- Toner, B.: The impact of agreement bias on the ranking of questionnaire response. *J. Soc. Psychol.* **127**, 221–222 (1987)
- van der Veld, W.M., Saris, W.E.: A unified model for the survey response process. Estimating the stability and crystallization of public opinion. In: Paper Presented at the European Association for Survey Research, Barcelona, July 18–21 (2005)
- Vermunt, J.K.: *Log-linear Models for Event Histories*. Sage Publications, Thousand Oakes, CA (1997)
- Viswanathan, M., Sudman, S., Johnson, M.: Maximum versus meaningful discrimination in scale response: implications for validity of measurement of consumer perceptions about products. *J. Bus. Res.* **57**, 108–124 (2004)